Protein Biomarkers for clinical application in colorectal cancer

de Wit, M.

2013

document version
Publisher's PDF, also known as Version of record

CHAPTER 4

COLORECTAL CANCER CANDIDATE BIOMARKERS IDENTIFIED BY TISSUE SECRETOME PROTEOME PROFILING

Submitted for publication

Meike de Wit, Huub Kant, Sander R. Piersma, Thang Pham, Sandra Mongera, Maaike P. A. van Berkel, Epie Boven, Fredrik Pontén, Gerrit A. Meijer, Connie R. Jimenez and Remond J. A. Fijneman.

ABSTRACT

Background: Colorectal cancer (CRC) is the second leading cause of cancer death in the Western world. Biomarkers associated with molecular changes in cancer cells can aid early detection, diagnosis, prognosis, therapy selection, and disease monitoring. Tumor tissue secretomes are a rich source of candidate biomarkers. While several secretome proteome datasets are available from pre-clinical CRC model systems, secretome data from clinically relevant human CRC and control tissues is lacking. The aim of this study was to identify CRC protein biomarkers that can be applied for development of laboratory tests to support clinical management of CRC.

Approach: Fresh human colon carcinoma tissue and patient-matched normal colon tissue samples were obtained from four CRC patients. Secretomes were collected by incubation in PBS and subjected to in-depth proteome profiling by a GeLC-MS/MS workflow. In addition, the secretomes of five CRC cell lines (i.e. HT-29, Caco-2, HCT116, SW480 and SW1398) were profiled. Quantitative comparisons were based on label-free spectral counting, *p*-values were calculated using the paired beta-binomial test, and further data analysis was performed using Ingenuity Pathway Analysis, SecretomeP, SignalP, STRING and Cytoscape software.

Results: A total of 2703 proteins were identified in tissue secretomes, of which 409 proteins were significantly more present in CRC samples than in controls. These proteins are associated with the biological processes RNA post-transcriptional modification, cell death and survival, cellular growth and proliferation, gene expression and protein synthesis. Biomarker selection was based on consistent and abundant protein over-representation in cancer-secretomes compared to control-secretomes, and presumed neoplastic origin based on overlap with the CRC cell line secretome proteome dataset. This procedure yielded 76 candidate human protein biomarkers. A subselection of biomarkers suited for early detection of CRC was revealed by overlap analysis with a previously obtained mouse model CRC secretome dataset, which revealed 21 candidates including proteins that have been proposed in literature as markers for CRC screening in stool or blood such as MCM5, TIMP-1 and LCN2.

Conclusion: Tissue secretome proteome profiling of human colon cancer is a powerful strategy to discover human candidate protein biomarkers for CRC. We identified 76 promising biomarker proteins that have potential to be used for development of a blood-based or stool-based test to support clinical management of CRC. Further studies in specified clinical settings are required to validate clinical applicability of these candidate biomarkers.

INTRODUCTION

Colorectal cancer is one of the most prevalent and deadly cancers, with an annual incidence of one million new cases[1]. The disease starts with the development of a non-malignant precursor lesion called an adenoma, of which only a minority of about 5% will eventually develop into CRC. Since there is such a well-defined precursor lesion, early detection of the disease seems the most realistic approach to reduce the currently high number of colorectal cancer deaths. Therefore, screening programs for detection of CRC are currently implemented in several countries. Yet, available methods for screening such as colonoscopy or detection of traces of blood in stool leave room for improvement[2,3]. Besides new screening biomarkers there is also a need for prognostic biomarkers, predictive biomarkers, and biomarkers that monitor treatment response and disease recurrence. Protein biomarkers that can be detected in blood or stool are of particular interest since they could be applied in a standard clinical setting alongside the routinely used markers such as CEA and CA19-9 (blood test) or haemoglobin (stool test)[4]. Identification of tumor-derived protein biomarkers in plasma is difficult since their concentration is five to eight orders of magnitude lower than the concentration of endogenous blood proteins such as albumin[5]. Tumor interstitial fluid proteins are derived from neoplastic cells and the surrounding microenvironment and enriched for proteins that are secreted, shed by membrane vesicles (a.o. exosomes), or externalized due to cell death[6]. Proteomic profiling of these fluids, further referred to as tumor secretomes, has led to the identification of novel candidate protein biomarkers for different tumor types[7-9].

The aim of the current study was to identify novel tumor-derived protein biomarkers that have the potential be applied in a blood-based or stool-based test for early diagnosis, prognosis, therapy prediction and/or disease monitoring of CRC.

MATERIAL AND METHODS

Patients

A total of four patients that underwent surgical resection at the VU University medical center (Amsterdam, the Netherlands) were included in this study. Collection, storage, and use of tissue and patient data were performed in accordance with the Code for Proper Secondary Use of Human Tissue in the Netherlands[10]. A pathologist inspected all samples to obtain information on tumor size, tumor and nodal stage, differentiation grade, and mucinous differentiation. For an overview of the clinicopathological characteristics, see table 1.

Tissue handling and tissue secretome collection

The tissue secretome collection was performed as described before[6]. In brief, following surgical resection, the specimen was immediately transferred to the pathology department where a pathologist excised a representative part of

Table 1 Clinicopathological characteristics of the colon cancer patients

Patient number[a]	Age	Gender	Tumor Size (CM)	Stage[b]	Tumor location	Differentiation grade	Mucinous	MSI/ MSS
1	92	F	5	C	Sigmoid colon	Moderate	no	MSS
2	80	F	3.5	B	Coecum	Moderate	no	MSS
3	78	M	1.3	A	Coecum	Moderate	no	MSS
4	61	M	10	C	Transversum	Moderate	yes	MSI

a This colomn represents an anonymization number code
b Dukes classification

the tumor and adjacent normal colon mucosa. Pieces of tissue were cut into cubes of approximately 1 mm^3 and rinsed in PBS to remove blood and stool particles. Subsequently the tissue cubes were incubated in 100µl PBS for 1 hour at 37°C. Next, samples were briefly centrifuged (2000 rpm at 4°C for 2 minutes) and the supernatant was transferred to a new eppendorf tube. The supernatants were centrifuged at maximum speed (13.200 rpm at 4°C for 20 minutes) to remove all remaining cells and debris. The soluble fractions further referred to as the 'tissue secretomes', were stored at -80°C until further use. The tissue cubes were processed by standard formalin fixation and paraffin embedding for histological evaluation. Microsatellite instability (MSI) status was determined using the MSI Analysis System version 1.2 (Promega Corporation, Madison, USA) as described previously[11].

Cell culture

The CRC cell lines HT-29 and Caco-2 were obtained from the American Type Culture Collection (LGC Standards GmbH, Wesel, Germany) and the HCT116, SW480 and SW1398 cell lines were kindly provided by Dr. G.J. Peters, Department of Oncology, VU University Medical Center. All cell lines were cultured in DMEM medium (Lonza Biowhittaker, Verviers, Belgium) containing 10% Fetal Bovine Serum (PAA, Pasching, Austria). Cell culture media were supplemented with 2 mM L-Glutamine, 100 IU/ml sodium-penicillin (Astellas Pharma B.V., Leiderdorp, The Netherlands) and 100 µg/ml streptomycin (Fisiopharma, Palomonta (SA), Italy).

Collection of cell line secretomes

Secretomes from cell lines were collected and processed as described before[12]. In brief, cells were cultured on 10 cm plates in normal complete medium containing FBS until 50-60% confluency was reached. Subsequently, the cells were washed four times in serum free medium and incubated for 24 hours in serum-free medium to obtain the secretome. The medium containing the secreted proteins was harvested by aspiration and passed through a 0.45 µm filter (Millex-HV, Millipore, Amsterdam, The Netherlands) to remove detached cells. The secretome was concentrated from 9 ML (3 plates) to 50 µL using a 10Kd MWCO centrifugal concentrator (Amicon ultra-4° C, Millipore,

Amsterdam, The Netherlands) operated at 4000 RPM and 4°C. Cell viability was assessed using Trypan blue stain (Lonza) and samples were only included in case of >90% living cells.

Gel electrophoresis and sample preparation for proteomics analysis

Protein concentrations were determined using the BCA protein assay (Pierce, Thermo Fisher Scientific, Rockford, USA). Twenty µg of proteins were separated by gel electrophoresis using a pre-cast 1D 4-12% gradient SDS-PAGE gel (Invitrogen, Carlsbad, USA). For gel images, see supplementary figures 2a and 3a. The gels were fixed in 50% ethanol containing 3% phosphoric acid and stained with Coomassie R-250. Gels were washed in MilliQ water and stored at 4°C until processing for in-gel digestion. Each lane was cut in 10 gel bands and further processed into tryptic peptides as described before[12]. The volume of the desalted peptide fractions was reduced to 50 µl in a vacuum centrifuge.

nanoLC-MS/MS proteomics analysis

Peptides were separated by an Ultimate 3000 nanoLC system (Dionex LC-Packings, Amsterdam, The Netherlands) equipped with a 20 cm × 75 µm ID fused silica column custom packed with 3 µm 120 Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany). After injection, peptides were trapped at 6 µl/min on a 1 cm × 100 µm ID trap column packed with 5 µm 120 Å ReproSil C18aqua at 2% buffer B (buffer A: 0.05% formic acid in MQ; buffer B: 80% acetonitrile + 0.05% formic acid in MQ) and separated at 300 nl/min in a 10–40% buffer B gradient in 60 min. Eluting peptides were ionized at 1.7 kV in a Nanomate Triversa Chip-based nanospray source using a Triversa LC coupler (Advion, Ithaca, NJ). Intact peptide mass spectra and fragmentation spectra were acquired on a LTQ-FT hybrid mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 50.000 in the ICR cell using a target value of 1×10^6 charges. In parallel, following an FT pre-scan, the top 5 peptide signals (charge-states 2+ and higher) were submitted to MS/MS in the linear ion trap (3 amu isolation width, 30 ms activation, 35% normalized activation energy, Q value of 0.25 and a threshold of 5000 counts). Dynamic exclusion was applied with a repeat count of 1 and an exclusion time of 30 seconds.

Database searching

To identify proteins from the acquired data MS/MS spectra were searched against the human IPI database v3.59 (80128 entries) using Sequest (version 27, rev 12), which is part of the BioWorks 3.3 data analysis package (Thermo Fisher, San Jose, CA). Following database searching the DTA and OUT files were imported into Scaffold Scaffold_2_06_01 (Proteome software, Portland, OR). Scaffold was used to organize the gel-band data and to validate peptide

identifications using the Peptide Prophet algorithm[13]. Only identifications with a probability >95% were retained. Subsequently, the Protein Prophet algorithm[14] was applied and protein identifications with a probability of >99% with 2 peptides or more in at least one of the samples were retained. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped.

Data mining and statistical analysis

For each protein identified, the number of assigned spectra was exported to Excel. Differential analysis of samples was performed using a paired Beta-Binominal test as described previously[15,16]. This modification resulted in a paired test thereby taking into account the origin of these samples e.g., comparing protein signatures between two tissues derived from the same patient. Additional general protein information was retrieved using Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com). Known and predicted protein-protein interactions were investigated using STRING version 9.0 (www.string-db.org)[17]. For cluster and gene ontology analyses we used the Cytoscape platform for network analysis (www.cytoscape.org)[18], using the plug-ins ClusterONE version 0.93 for clustering and BINGO version 2.44 for the analysis of biological processes within our obtained networks based on GO annotations[19,20]. Additionally, transmembrane domains and signal peptide sequences were investigated using secretomeP server 2.0[21].

MCM5 immunohistochemical staining of tissue microarrays

Tissue MicroArrays (TMAs) had been constructed previously[11] using a series of 82 colon adenomas and 82 CRCs that were collected retrospectively from 2001 to 2008 at the VU University Medical Center, Amsterdam, The Netherlands. Collection, storage, and use of tissue and patient data were performed in accordance with the Code for Proper Secondary Use of Human Tissue in the Netherlands[10]. Two to six cores from each paraffin block were included in the TMAs, depending on the amount of tissue available. TMA sections were deparaffinized and rehydrated and endogenous peroxidases were blocked with 0.3% hydrogen peroxide in methanol. Antigens were retrieved by autoclaving in 10mM citrate buffer solution (pH 6.0). Primary mouse anti-MCM5 antibody (HPA00845, atlas antibodies, Stockholm, Sweden) was incubated at a 1:1000 dilution in antibody dilutend (DAKO) at 4°C for 30 minutes, and subsequently detected by a HRP-coupled anti-mouse polymer (Envision, DAKO, Heverlee, Belgium) followed by incubation with Diaminobenzidine (DAKO). Sections were counterstained with Mayer's haematoxylin.

The stained TMA sections were automatically scanned with a digital pathology system Aperio ScanScope XT Slide Scanner (Aperio Technologies, Vista, California, USA) equipped with a 20x objective with a numerical aperture

of 0.75 (from Olympus, Tokyo, Japan), and stainings were scored using the ImageScope viewer (Aperio). The scoring methods applied were guided by the staining pattern observed. The percentage of nuclear staining in the epithelium was scored into 5 categories, 0-1%, 2-10%, 11-25%, 26-50%, 51-75% or more than 75% staining among all nuclei in a tissue core. The highest score of multiple cores from each tumor was used for further analysis. Final evaluation was performed by dichotomizing the data into less than 75% or more than 75% of positive nuclei. Due to loss of cores during the staining procedure, i.e. for technical reasons, stainings could be evaluated for 62 colon adenomas and 58 CRCs.

RESULTS

Identification of proteins in CRC and normal colon tissue secretomes

Tissue secretomes were collected from four CRC-normal colon tissue pairs and processed for proteome analysis by mass spectrometry. Patient and tumor characteristics are provided by table 1, and an overview of the workflow by supplementary figure 1. A total of 2703 non-redundant proteins were identified, with an average of 1989 proteins per cancer- or normal colon secretome sample (ranging from 1264 to 2292 proteins). Out of these 2703 proteins, 2366 were identified both in CRC secretomes and in normal colon tissue secretomes, 283 only in the CRC secretomes and 54 only in the normal tissue secretomes. The number of identified proteins was higher in the CRC secretomes than in the normal tissue secretomes (on average 2198 and 1780, respectively). In total, 961 out of 2420 (40%) of proteins were identified in all four normal colon tissue secretomes and 1627 out of 2649 (61%) were detected in all four cancer secretomes (supplementary figures 2b and 2c). A complete overview of all identified proteins is provided by supplementary table 1.

Comparison of CRC to normal colon tissue secretomes

To obtain an overview of the tissue secretome dataset, an unsupervised hierarchical clustering was performed using the normalized spectral count data from all 2703 identified proteins (figure 1a). The normal tissue secretome sample of patient 3 clustered apart from all other samples, which may be explained by the relatively low amount of proteins identified in this one particular sample (n=1264) compared to the other 7 samples (average n=2089). The cluster of 7 samples exhibited two sub-clusters, one containing the three remaining normal tissue samples and the other all four cancer samples. The cancer secretome cluster separated patients 1-3 from patient 4, thereby separating the three microsatellite stable tumors from the one MSI tumor (table 1). Importantly, the unsupervised cluster analysis indicated that the secretome protein pattern was more similar among the four cancer secretome samples than among patient-matched cancer-normal colon tissue pairs.

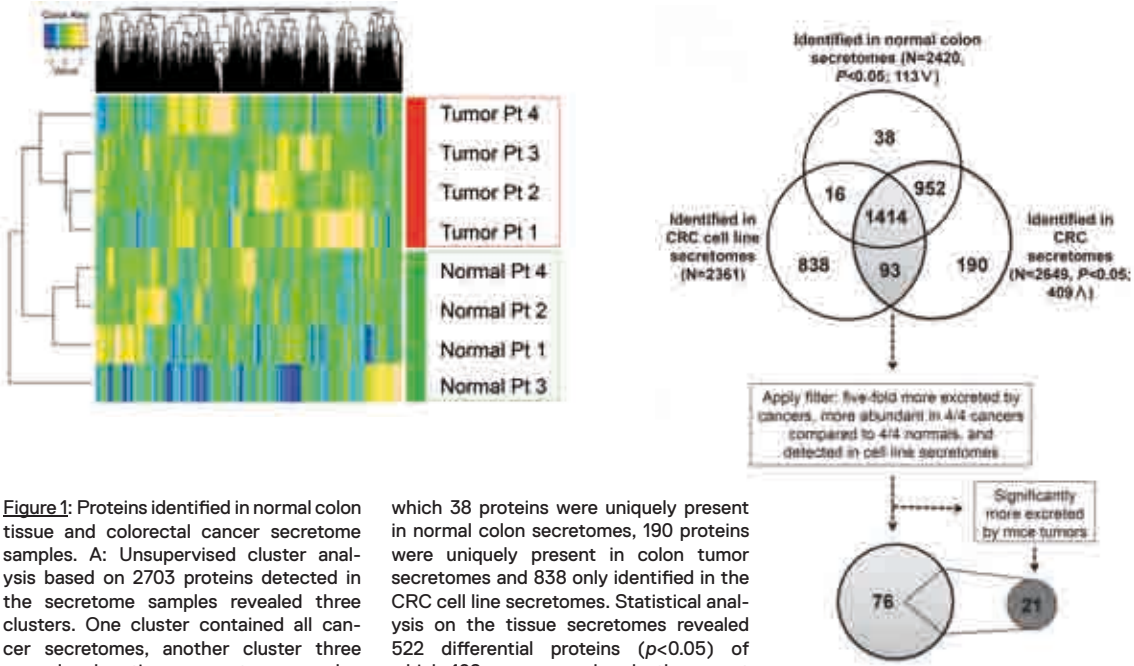


Figure 1: Proteins identified in normal colon tissue and colorectal cancer secretome samples. A: Unsupervised cluster analysis based on 2703 proteins detected in the secretome samples revealed three clusters. One cluster contained all cancer secretomes, another cluster three normal colon tissue secretomes, and a third cluster the remaining normal colon tissue secretome. The range of relative protein abundances is indicated from low (blue) to high (orange). B: The venn diagram illustrates overlapping and unique protein identifications in the four normal colon tissue secretome samples, the four paired cancer secretome samples, and five CRC cell line secretomes. A total of 3541 different proteins were identified, of which 38 proteins were uniquely present in normal colon secretomes, 190 proteins were uniquely present in colon tumor secretomes and 838 only identified in the CRC cell line secretomes. Statistical analysis on the tissue secretomes revealed 522 differential proteins ($p<0.05$) of which 409 were more abundantly present in cancer secretomes and 133 in normal tissue secretomes. Applying a filter for the selection of top candidate biomarkers led to the selection of 76 proteins. Overlap analysis with a previously obtained dataset for proteins that are significantly more abundantly excreted by early-stage mouse colon tumors revealed 21 proteins, which represent candidate-biomarkers for early detection of CRC.

Candidate CRC biomarkers should be more abundant in tumor secretomes compared to the normal colon tissue secretomes. To identify such proteins the complete protein dataset was further analyzed with a statistical test for paired samples, i.e. taking into account the relationship between normal and cancer tissue samples from individual patients, to identify CRC associated proteins[16]. From the 2703 proteins, 409 proteins were significantly more abundantly present in CRC secretomes ($p<0.05$) and 113 proteins were less abundant in CRC secretomes compared to normal tissue secretomes (supplementary table 1).

CRC cell line secretome proteins and comparison to tissue secretomes

Proteins identified in normal colon tissue- and cancer-secretomes may be derived from epithelial cells, from the surrounding microenvironment i.e. stromal cells and extracellular matrix, or from blood. In order to determine which proteins may originate from neoplastic epithelial cells secretomes from five CRC cell lines (HT-29, Caco-2, HCT116, SW480 and SW1398) were

collected. These were processed and analyzed by nanoLC-MS/MS similar to the tissue secretomes (supplementary figures 1 and 3a). In total 2361 proteins were identified in these cell line secretomes, for an overview see supplementary table 2. Almost half of these proteins (n=1158, 49%) were commonly identified in all five cell line secretomes (supplementary figure 3b).

In total 56% (1523 of 2703) of the proteins identified in the tissue secretomes were also present in these cell line secretomes (figure 1b and supplementary table 1). These proteins are more likely to be derived from the colon cancer cells than from the surrounding tumor stroma.

General CRC tissue and cell line secretome protein characteristics

To gain more insight in the processes and pathways in which CRC secretome proteins are involved, pathway analysis was performed using IPA (Ingenuity® Systems, www.ingenuity.com). First, the top molecular and cellular functions associated with the significantly differentially identified proteins were analyzed. For the proteins that were more abundant in CRC secretomes, the top function was RNA post-transcriptional modification. For the proteins that were more abundant in the normal colon tissue secretomes this was cellular assembly and organization (data not shown).

Secondly, the subcellular localization of the secretome proteins was retrieved from the IPA database. Overall, the proportion of proteins from the different subcellular origins was comparable between the normal colon tissue and cancer tissue secretome samples (figures 4a and 4b). The subcellular origins of the CRC cell line secretome proteins were similarly distributed as the normal colon and cancer tissue secretomes (figure 4c). Remarkably, among the proteins significantly more abundant in the CRC secretomes, the proportion of nuclear proteins (40%) was much higher than that among the proteins less abundant in CRC secretomes (5%) (figures 4d and 4e).

Mechanisms of protein secretion

Proteins present in the tissue secretomes could be actively secreted, shed by membrane vesicles (e.g. exosomes), or externalized due to cell death. To investigate if the identified proteins were actively secreted the SignalP server was used for prediction of secretion by the classical pathway, i.e. in an endoplasmic reticulum/Golgi dependent secretory pathway depending on presence of a signal peptide, and the SecretomeP server for prediction of secretion via the non-classical pathways, e.g. through membrane transporters or exosomes[21]. Almost half of the 2669 proteins (1285; 48%) for which a result was obtained were predicted to be secreted, of which 440 (34%) via the classical pathway and 845 (66%) via the non-classical pathway. Among the proteins that were differentially identified between CRC and normal colon tissue secretomes, the ones that were more abundant in the normal tissue secretomes were in

general more often predicted to be secreted (n=64, 59%) than the proteins more abundant in the CRC secretomes (n=147, 36%) (supplementary table 1). On the other hand, the cancer secretomes contained a higher proportion of non-classically secreted proteins (n=103, 70%) than the normal colon tissue secretomes (n=19, 30%; $p=5.2 \times 10^{-8}$).

To estimate to what extent proteins might be excreted by tumor cells through vesicular transport as microvesicle cargo, the tissue secretome protein data were compared to a list of proteins previously identified by in-depth GeLC-MS/MS proteomics analysis of the microvesicle fraction, the soluble fraction and the complete cell lysate of the human HT-29 CRC cell line secretome (supplementary table 1)[7]. Out of 2085 proteins that were both found in the HT-29 dataset and the tissue secretomes, 349 proteins were significantly more abundant in the cancer tissue secretomes and 65 proteins more abundant in the normal colon tissue secretomes. Among these, 50% of the cancer-enriched proteins were found in microvesicles (174 out of 349) compared to 25% of proteins enriched in normal tissue secretomes (16 out of 65). These data show that the tumor secretomes were significantly enriched for microvesicle related proteins (χ^2: $p=2.1 \times 10^{-4}$).

Out of the 349 proteins that were more abundant in cancer secretomes and detected in the HT-29 dataset, 101 proteins were annotated as nuclear. Of these, 75 (74%) were found in the microvesicle fraction while from the 248 non-nuclear proteins only 99 (40%) were detected in microvesicles. This proportion of nuclear proteins in the microvesicle fraction was significantly larger than the proportion of non-nuclear proteins in this fraction (supplementary table 1, χ^2: $p=4.8 \times 10^{-9}$).

Selection of promising candidate biomarkers for distinct clinical applications

The list of identified proteins can be mined for selection of promising CRC candidate biomarkers that can be detected in blood or stool for various clinical applications such as early detection, diagnosis, prognosis, therapy response prediction, and disease monitoring of CRC patients. To narrow down the list of putative biomarkers the following criteria were applied: (1) CRC is a heterogeneous tumor and therefore only proteins that were consistently (in all four sample pairs) more abundant in the tumor secretome compared to the paired normal tissue secretome were selected; (2) candidate biomarkers should be highly increased in blood or stool of CRC patients compared to healthy individuals, and for that reason a cutoff of at least five-fold enrichment in tumor secretomes compared to normal tissue secretomes was applied; (3) to focus on proteins that are specific for neoplastic cells only proteins that were also detected in the CRC cell line secretomes were included. Applying these criteria led to the selection of 76 promising secretome based candidate CRC biomarkers (figure 1b and table 2).

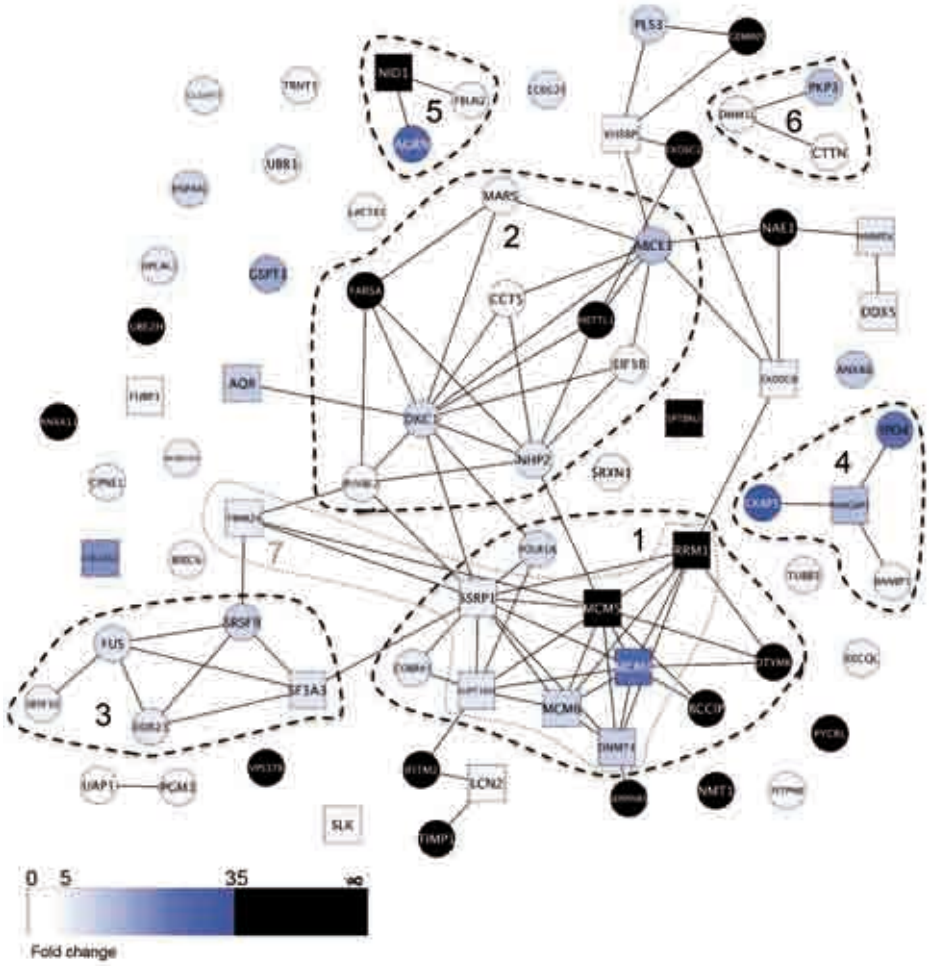


Figure 2: Protein network of 76 top-candidate CRC biomarkers. Nodes represent interacting proteins, the lines between them represent known and predicted interactions. Color intensity indicates fold changes in CRC compared to normal colon tissue secretomes, from low (white) to high (black). The dashed lines indicate sub-clusters which are linked to specific molecular functions: 1. DNA replication; 2. RNA metabolic process; 3. RNA splicing; 4. cytoskeleton polarity; 5. extracellular structure organization; 6. mitochondrial fragmentation. Squares indicate the 21 proteins that overlap with the mouse colon tumor study and represent candidate biomarkers for early detection. The grey dotted line indicates the single sub-cluster identified among these 21 early detection candidate biomarkers, associated to DNA replication.

To obtain proof of concept, a subselection in this dataset was made for proteins that could be applied for early detection of CRC. To this end an overlap analysis was performed with proteins identified in a previous comparative study on mouse colon tumor secretomes and normal colon tissue secretomes[7]. These mouse tumors represent relative early stage CRC and proteins that are differentially secreted by those tissues would therefore be promising biomarkers for a blood-based or stool-based test for early detection of CRC. Out of a total of 2172 proteins identified in the mouse study, 1582 proteins overlapped with the current human study (supplementary figure 4a and supplementary table 1). Of these 1582 proteins 318 were significantly more excreted by mouse colon

tumors compared to normal mouse colon tissue ($p<0.05$), and 409 proteins were significantly more excreted by the human CRCs compared to their matched normal colon tissue counterparts ($p<0.05$). Among these candidate biomarkers, 90 proteins were shared between the mouse and human differential datasets (supplementary figure 4b), including 21 candidate biomarkers that also fulfilled the additional selection criteria of consistent (in four cancer-normal pairs) and abundant (>5-fold) excretion by the human CRCs (figure 1b and table 2). These 21 proteins therefore represent promising candidate biomarkers for early detection of CRC.

Pathway analysis of candidate biomarkers

To further identify the underlying protein complexes we employed the Cytoscape tool[18]. The top 76 candidate biomarker proteins were imported into the STRING tool[17] to generate a protein-protein interaction map of known and predicted interactions between these proteins. The protein-protein interaction map was imported into CytoScape software upon which six significant sub-clusters (i.e. connections within the cluster that were stronger than those with molecules on the outside, $p<0.05$) were extracted using the ClusterOne tool. Next, gene ontology analysis was performed with the BINGO tool (figure 2). For each sub-cluster, gene ontology annotations for biological processes were assigned. Relative enrichment of biological functions was calculated by comparing the number of proteins of a sub-cluster involved in that specific biological process to the number of involved proteins from the entire proteome (supplementary table 2). The proteins within sub-cluster one were found to be associated with DNA replication, sub-cluster two was linked to RNA metabolic processes while sub-cluster three was linked to RNA splicing. Sub-cluster four was associated with cytoskeleton polarity; sub-cluster five was linked to extracellular structure organization and sub-cluster six to mitochondrial fragmentation and apoptosis. A separate analysis on the 21 candidate biomarkers for early detection of CRC (indicated by squares in figure 2) revealed a single cluster containing 7 proteins of sub-cluster one, i.e. SSRP1, SUPT16H, RRM1, DNMT1, MCM3, MCM5, and MCM6 combined with the TRIM28 protein (figure 2). This sub-cluster was also associated to DNA replication, similar to the complete sub-cluster one.

Remarkably, six members of the minichromosome maintenance (MCM) family were identified and five of those were significantly more excreted by the cancers (supplementary table 1). All of these proteins were also significantly more excreted by the mouse colon tumors, indicating that they may be of interest for early detection. From this family the MCM5 protein was the protein with most protein-protein interactions (n=10) and the only one that was uniquely detected in cancer secretomes. Immunohistochemical staining for MCM5 was performed on tissue micro arrays containing colon adenomas (i.e. precursor lesions of CRC) and carcinomas (supplementary table 4 for details) to assess potential overexpression in CRC. MCM5 was expressed by

all tissues, however, the percentage of cells with nuclear staining varied among the different samples (figures 3a and 3b). When the data was dichotomized into a 'high expressing' group with >75% of all epithelial cells showing nuclear staining *versus* a 'low expressing' group where the expression was <75% of the epithelial cells, the CRCs showed significantly more frequently the high expression pattern than the adenomas (p=0.03; figure 3c). These data indicate that MCM5 is ubiquitously expressed and is overexpressed by CRC.

DISCUSSION

The present study aimed to identify novel CRC protein biomarkers that can be applied for development of blood-based or stool-based CRC tests. In-depth proteomics analysis of tissue secretomes revealed candidate biomarkers that were excreted by cancer tissues. We report the identification of 2703 proteins in total and 76 promising CRC candidate biomarkers that were selected based on stringent criteria (consistent and abundant excretion by CRC tissue and presumably of neoplastic cell origin). These 76 proteins are the most promising candidate biomarkers for detection in a blood-based or stool-based test. The

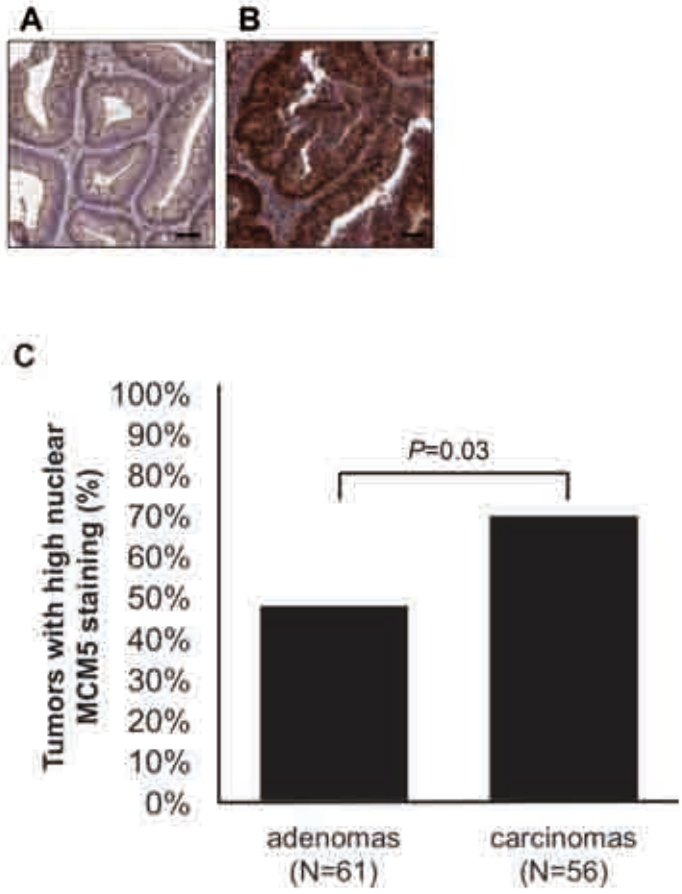


Figure 3: MCM5 immunohistochemical staining of colon adenomas and carcinomas. TMAs containing a series of colon adenomas and cancers were stained for MCM5 by immunohistochemistry. Representative examples show that MCM5 staining of epithelial cells in adenomas (A) was less intense than that in carcinomas (B). Epithelial staining was mostly nuclear. Scale bars indicate 50 µm. A: adenoma with low MCM5 staining. B: Carcinoma with high nuclear MCM5 staining (>75% positive nuclei). C: Quantification of MCM5 immunohistochemical staining of TMAs reveals that high levels of MCM5 staining were more frequently detected in carcinomas (67.2%) compared to adenomas (46.8%).

Table 2: List of 76 CRC top-candidate biomarkers

Information obtained from IPA					Beta-binomial paired test		Raw spectral counts					
Accession Number	Gene Symbol	Entrez Gene Namew	Subcellular Location	Function	Fold Change	P-value	Pt. 1 tumor	Pt. 2 tumor	Pt. 3 tumor	Pt. 4 tumor	HT-29 Microvesicles	Mouse tumor secretome vs. normal secretome study*#
IPI00303207	*ABCE1*	ATP-binding cassette, sub-family E (OABP), member 1	Cytoplasm	transporter	13,77	2,10E-03	8	9	1	2	10	down
IPI00374563	*AGRN*	agrin	Plasma Membrane	other	28,44	5,06E-07	28	12	21	20	45	-
IPI00221234	*ALDH7A1*	aldehyde dehydrogenase 7 family, member A1	Cytoplasm	enzyme	6,63	1,05E-02	3	7	6	1	0	down
IPI00414320	*ANXA11*	annexin A11	Nucleus	other	∞	5,04E-03	1	1	4	2	13	not found
IPI00221226	*ANXA6*	annexin A6	Plasma Membrane	other	12,12	1,31E-05	15	40	18	30	0	not found
IPI00297572	*AQR*	aquarius homolog (mouse)	Nucleus	other	11,67	1,40E-03	2	3	1	20	3	up
IPI00002203	*BCCIP*	BRCA2 and CDKN1A interacting protein	Nucleus	other	∞	9,81E-03	2	1	1	2	0	not found
IPI00299635	*BIRC6*	baculoviral IAP repeat containing 6	Cytoplasm	enzyme	5,70	6,90E-03	14	6	1	6	0	-
IPI00410091	*C11orf73*	chromosome 11 open reading frame 73	Cytoplasm	other	6,85	3,66E-02	1	4	1	2	0	not found
IPI00396174	*CCDC25*	coiled-coil domain containing 25	unknown	other	8,10	1,85E-02	1	4	3	2	0	not found
IPI00010720	*CCT5*	chaperonin containing TCP1, subunit 5 (epsilon)	Cytoplasm	other	5,46	5,82E-04	19	19	8	9	7	-
IPI00028275	*CKAP5*	cytoskeleton associated protein 5	Nucleus	other	32,52	1,38E-05	14	9	4	13	0	-
IPI00103483	*COBRA1*	cofactor of BRCA1	Nucleus	other	7,17	2,24E-02	1	4	3	1	0	not found
IPI00329573	*COL12A1*	collagen, type XII, alpha 1	Extracellular Space	other	20,05	6,33E-08	48	65	33	147	0	down
IPI00018452	*CPNE1*	copine I	unknown	transporter	6,15	3,22E-03	8	6	6	4	5	not found
IPI00029601	*CTTN*	cortactin	Plasma Membrane	other	5,38	5,32E-03	8	3	5	5	6	-
IPI00006725	*DDX23*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	Nucleus	enzyme	7,46	1,91E-02	1	4	2	2	0	not found
IPI00017617	*DDX5*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	Nucleus	enzyme	6,48	4,03E-04	15	17	10	7	19	up
IPI00037283	*DNM1L*	dynamin 1-like	Cytoplasm	enzyme	5,64	2,33E-03	17	10	12	5	0	down
IPI00031519	*DNMT1*	DNA (cytosine-5-)-methyltransferase 1	Nucleus	enzyme	11,55	5,70E-04	6	3	2	16	2	up
IPI00013862	*DTYMK*	deoxythymidylate kinase (thymidylate kinase)	Cytoplasm	kinase	∞	2,32E-03	3	2	1	3	5	not found
IPI00299254	*EIF5B*	eukaryotic translation initiation factor 5B	Cytoplasm	translation regulator	5,42	4,79E-03	16	6	5	7	5	-
IPI00015905	*EXOSC2*	exosome component 2	Nucleus	enzyme	∞	9,54E-04	2	3	2	4	0	-
IPI00552920	*EXOSC8*	exosome component 8	Nucleus	enzyme	6,39	4,20E-02	2	1	2	3	0	up
IPI00031820	*FARSA*	phenylalanyl-tRNA synthetase, alpha subunit	Cytoplasm	enzyme	∞	2,60E-04	10	3	2	2	4	-

Table 2: List of 76 CRC top-candidate biomarkers (continued)

Information obtained from IPA					Beta-binomial paired test		Raw spectral counts					
Accession Number	Gene Symbol	Entrez Gene Namew	Subcellular Location	Function	Fold Change	P-value	Pt. 1 tumor	Pt. 2 tumor	Pt. 3 tumor	Pt. 4 tumor	HT-29 Microvesicles	Mouse tumor secretome vs. normal secretome study*#
IPI00023824	*FBLN2*	fibulin 2	Extra-cellular Space	other	5,75	2,86E-03	10	12	1	5	na	not found
IPI00375441	*FUBP1*	far upstream element (FUSE) binding pro-tein 1	Nucleus	transcription regulator	5,36	4,68E-03	14	10	5	5	6	up
IPI00221354	*FUS*	fused in sarcoma	Nucleus	transcription regulator	8,42	2,57E-04	13	14	11	13	4	-
IPI00291783	*GEMIN5*	gem (nuclear orga-nelle) associated protein 5	Nucleus	other	∞	9,81E-03	2	1	1	2	0	not found
IPI00909083	*GSPT1*	G1 to S phase transi-tion 1	Cytoplasm	translation regulator	16,56	1,10E-03	10	6	2	3	0	-
IPI00011274	*HNRPDL*	heterogeneous nuclear ribonucleoprotein D-like	Nucleus	other	7,23	1,41E-03	12	5	5	6	6	up
IPI00219344	*HPCAL1*	hippocalcin-like 1	Cytoplasm	other	7,10	2,58E-02	2	2	2	3	2	not found
IPI00295485	*HSPA4L*	heat shock 70kDa pro-tein 4-like	Cytoplasm	other	10,10	1,11E-02	11	1	1	1	0	-
IPI00008922	*IFITM2*	interferon induced transmembrane protein 2 (1-8D)	Plasma Membrane	other	∞	2,38E-03	1	3	1	4	0	not found
IPI00156374	*IPO4*	importin 4	Nucleus	transporter	24,93	4,93E-05	17	6	5	5	4	-
IPI00479786	*KHSRP*	KH-type splicing regu-latory protein	Nucleus	enzyme	5,95	1,10E-03	22	15	6	15	5	up
IPI000052	*LACTB2*	lactamase, beta 2	Cytoplasm	other	5,31	1,64E-02	4	3	2	4	0	-
IPI00299547	*LCN2*	lipocalin 2	Extra-cellular Space	transporter	5,58	2,92E-04	23	4	90	48	4	up
IPI00008240	*MARS*	methionyl-tRNA syn-thetase	Cytoplasm	enzyme	5,57	2,64E-03	17	10	6	9	12	-
IPI00018350	*MCM5*	minichromosome main-tenance complex compo-nent 5	Nucleus	enzyme	∞	4,54E-05	12	6	1	4	9	up
IPI00031517	*MCM6*	minichromosome main-tenance complex compo-nent 6	Nucleus	enzyme	10,66	9,73E-04	11	6	1	8	11	up
IPI00290184	*METTL1*	methyltransferase like 1	Nucleus	enzyme	∞	6,56E-03	3	2	1	1	0	not found
IPI00018968	*NAE1*	NEDD8 activating enzyme E1 subunit 1	Cytoplasm	enzyme	∞	1,21E-04	5	8	3	2	0	-
IPI00041325	*NHP2*	NHP2 ribonucleoprotein homolog (yeast)	Nucleus	other	7,91	1,65E-02	2	3	3	2	2	not found
IPI00026944	*NID1*	nidogen 1	Extra-cellular Space	other	∞	7,37E-04	4	2	3	3	0	up
IPI00329692	*NMT1*	N-myristoyltrans-ferase 1	Cytoplasm	enzyme	∞	2,32E-03	3	2	1	3	0	-
IPI00030116	*PGM3*	phosphoglucomutase 3	Cytoplasm	enzyme	5,69	3,74E-03	11	7	6	5	0	-
IPI00334907	*PITPNB*	phosphatidylinositol transfer protein, beta	Cytoplasm	transporter	5,44	1,27E-03	11	9	7	11	0	-
IPI00026952	*PKP3*	plakophilin 3	Plasma Membrane	other	13,25	1,87E-03	15	3	5	1	11	-

Table 2: List of 76 CRC top-candidate biomarkers (continued)

Information obtained from IPA					Beta-binomial paired test		Raw spectral counts					
Accession Number	Gene Symbol	Entrez Gene Namew	Subcellular Location	Function	Fold Change	P-value	Pt. 1 tumor	Pt. 2 tumor	Pt. 3 tumor	Pt. 4 tumor	HT-29 Microvesicles	Mouse tumor secretome vs. normal secretome study*#
IPI00216694	*PLS3*	plastin 3	Cytoplasm	other	10,48	9,90E-04	11	10	5	1	4	-
IPI00031960	*POLR1A*	polymerase (RNA) I polypeptide A, 194kDa	Nucleus	enzyme	9,36	1,84E-03	6	2	7	9	0	-
IPI00646105	*PYCRL*	pyrroline-5-carboxy-late reductase-like	unknown	enzyme	∞	5,67E-03	1	3	1	2	0	-
IPI00414127	RANBP1	RAN binding protein 1	Nucleus	other	5,64	8,50E-03	13	3	2	4	3	-
IPI00294879	RANGAP1	Ran GTPase activating protein 1	Cytoplasm	other	14,23	2,25E-03	8	6	4	1	2	up
IPI00178431	RECQL	RecQ protein-like (DNA helicase Q1-like)	Nucleus	enzyme	6,32	5,15E-03	11	5	4	5	4	not found
IPI00013871	RRM1	ribonucleotide reduc-tase M1	Nucleus	enzyme	∞	7,81E-04	6	1	1	5	4	up
IPI00009104	RUVBL2	RuvB-like 2 (E. coli)	Nucleus	transcription regulator	5,64	3,00E-02	5	6	1	1	18	not found
IPI00783625	SER-PINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	Extra-cellular Space	other	∞	1,03E-07	19	10	10	12	2	-
IPI00022827	SLK	STE20-like kinase	Nucleus	kinase	5,71	2,54E-03	12	5	3	14	0	up
IPI00012645	SPTBN2	spectrin, beta, non-erythrocytic 2	Cytoplasm	other	∞	6,67E-04	8	3	2	1	0	up
IPI00009071	SRSF10	serine/arginine-rich splicing factor 10	Nucleus	other	6,59	2,44E-03	7	8	3	6	0	not found
IPI00012340	SRSF9	serine/arginine-rich splicing factor 9	Nucleus	enzyme	10,60	6,49E-04	5	7	5	9	2	not found
IPI00168554	SRXN1	sulfiredoxin 1	Cytoplasm	enzyme	6,00	3,70E-02	3	1	2	2	0	not found
IPI00005154	SSRP1	structure specific rec-ognition protein 1	Nucleus	other	6,77	3,49E-03	6	5	2	11	6	up
IPI00026970	SUPT16H	suppressor of Ty 16 homolog (S. cerevisiae)	Nucleus	transcription regulator	8,10	2,77E-04	15	12	8	21	14	up
IPI00032292	TIMP1	TIMP metallopeptidase inhibitor 1	Extra-cellular Space	other	∞	7,20E-04	4	2	2	4	3	not found
IPI00438229	TRIM28	tripartite motif con-taining 28	Nucleus	transcription regulator	7,04	1,84E-04	20	18	9	18	8	up
IPI00289807	TRNT1	tRNA nucleotidyl trans-ferase, CCA-adding, 1	Cytoplasm	enzyme	5,80	1,75E-02	3	3	5	3	0	-
IPI00013683	TUBB3	tubulin, beta 3	Cytoplasm	other	6,56	5,50E-03	8	3	5	2	0	down
IPI00000684	UAP1	UDP-N-acteylglu-cosamine pyrophospho-rylase 1	Nucleus	enzyme	5,04	1,10E-02	4	10	2	2	2	down
IPI00020965	UBE2H	ubiquitin-conjugating enzyme E2H	Cytoplasm	enzyme	∞	9,97E-03	1	1	1	3	0	not found
IPI00217405	UBR1	ubiquitin protein ligase E3 component n-recognin 1	Cytoplasm	enzyme	6,82	7,50E-03	9	4	1	7	0	not found
IPI00002926	VPS37B	vacuolar protein sort-ing 37 homolog B (S. cerevisiae)	Cytoplasm	other	∞	1,76E-03	3	1	2	4	2	not found

Fijneman RJA, de wit M, Pourghiasian M, Piersma SR, Pham TV, Warmoes MO, et al. Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diag-nosis of human colorectal cancer. Clin Cancer Res 2012; 18:2613–2624.

*∧ more or ∨ less; secreted by mouse tumors than normal tissues, na; not detected, -; no difference in secretion.

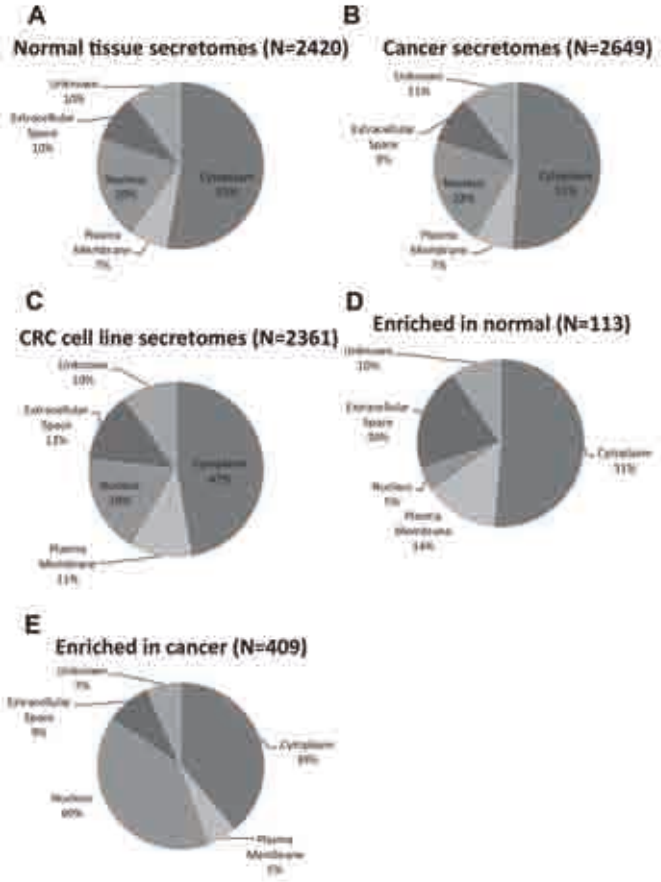


Figure 4: Proportion of protein subcellular localizations. The pie charts show the percentage of different subcellular localizations of the identified proteins. A: proteins identified in normal tissue secretomes, B: proteins identified in cancer secretomes, C: proteins identified in CRC cell line secretomes, D: proteins significantly less abundant in CRC secretomes, E: proteins significantly more abundant in CRC secretomes.

clinical setting in which these biomarkers may be useful needs to be established by other experiments. Proof of concept was provided by overlap analysis with a previously obtained mouse model for early stage CRC dataset, which yielded 21 promising candidate biomarkers for early detection. The proteins that were more abundant in the cancer secretomes comprised many nuclear proteins functioning in DNA replication, cell division and other processes commonly associated with cancer. One of the proteins involved in DNA replication, MCM5, was confirmed as being overexpressed by CRCs. In addition, strong indications that a large proportion of these nuclear proteins may be externalized through microvesicle-mediated secretion were found by integrating the tissue secretome data with a previously generated HT-29 cell line microvesicle dataset.

Minichromosome maintenance (MCM) proteins are essential for DNA replication, which are present in all cells whilst in active cell cycle, while during differentiation these proteins are downregulated[22]. This feature results in positive staining for dividing cells, which in normal colon are restricted to the base of the crypt. Besides MCM5 this study also identified MCM2-MCM7, of which MCM3 and MCM6 were also part of the list of 21 candidate biomarkers for

early detection. The MCM proteins have previously been linked to CRC and were detected in whole CRC cells isolated from stool of (early stage) CRC patients[22]. Therefore, these proteins show high potential to be applied in a stool-based CRC test. Plasma tissue inhibitor of metalloproteinases-1 (TIMP-1) is another proposed candidate marker for early detection, as proteins levels in both blood and stool are increased in CRC patients[2]. High specificity levels (>95%) have been found for both CRC as well as advanced adenomas, and a recent study on plasma samples of 4509 colonoscopy controlled individuals suggests that TIMP-1 in combination with CEA could aid in early detection of CRC[23]. In addition to these potential biomarkers for early diagnosis also proteins linked to disease progression and disease prognosis were identified. DNMT1 is a DNA-methyltransferase that regulates gene expression via promotor methylation that is overexpressed in CRC; aberrant methylation patterns are commonly found in CRC[24,25]. LCN2 (also known as NGAL) has previously been shown to be overexpressed during adenoma–to-carcinoma progression and to be associated with stage, tumor recurrence and overall survival. Moreover, plasma levels of LCN2 were found to be higher in more advanced stages than earlier stages, indicating that this protein may serve as a disease monitoring or treatment response biomarker[26]. SerpinB5 (also known as Maspin) was associated to worse disease outcome in stage III CRC patients (N. Snoeren, dept. of surgery UMC Utrecht, unpublished data). In addition, SerpinB5 transcript levels were found to be elevated in peripheral blood of CRC patients compared to controls[27]. The ABCE1 protein is a member of the ATP-binding cassette transporter family that can transport chemotherapeutic agents out of a cell, thereby facilitating drug resistance. ABCE1 mRNA levels were also previously found to be upregulated in tumor *versus* control tissue[28]. In summary, the CRC candidate biomarkers here identified include multiple proteins with known potential relevance for CRC tumor biology that await application for a clinical purpose.

The pathway analysis of CRC candidate biomarkers revealed several functional nuclear protein complexes to be secreted, such as proteins involved in DNA replication and repair, and several members of the splicing machinery. In fact, one of the most remarkable findings of this work is the proportion of nuclear proteins among the more abundantly *versus* the less abundantly excreted proteins by cancer tissues compared to normal colon tissue secretomes (40% vs. 5%, respectively). This observation is similar to the results from mouse colon (tumor) secretomes, in which proteins significantly more abundant in tumor samples contained 42% nuclear proteins while the proteins enriched in normal tissue secretomes contained only 6% nuclear proteins[7]. Secretion of proteins that lack a signal peptide, such as nuclear proteins, may be mediated via microvesicles. Increased microvesicle secretion is an emerging tumor hallmark, and the amount of studies showing an active role for such microvesicles in tumor development as well as tumor progression

is rapidly growing[29-31]. In line with this, the amount of non-classically secreted proteins was increased in the tumor secretomes compared to the normal tissue secretomes. Nevertheless, not all nuclear proteins were annotated as non-classically secreted upon prediction by SecretomeP software. An overlap analysis was performed with proteins previously identified in the microvesicle fraction of HT-29 cells, and indeed many nuclear proteins were identified in this microvesicle fraction. The proportion of nuclear proteins in microvesicles was significantly larger than the proportion of non-nuclear proteins found in the microvesicles. Therefore, these data imply large-scale microvesicle-mediated nuclear protein secretion, which exceeds the current prediction of non-classical secretion.

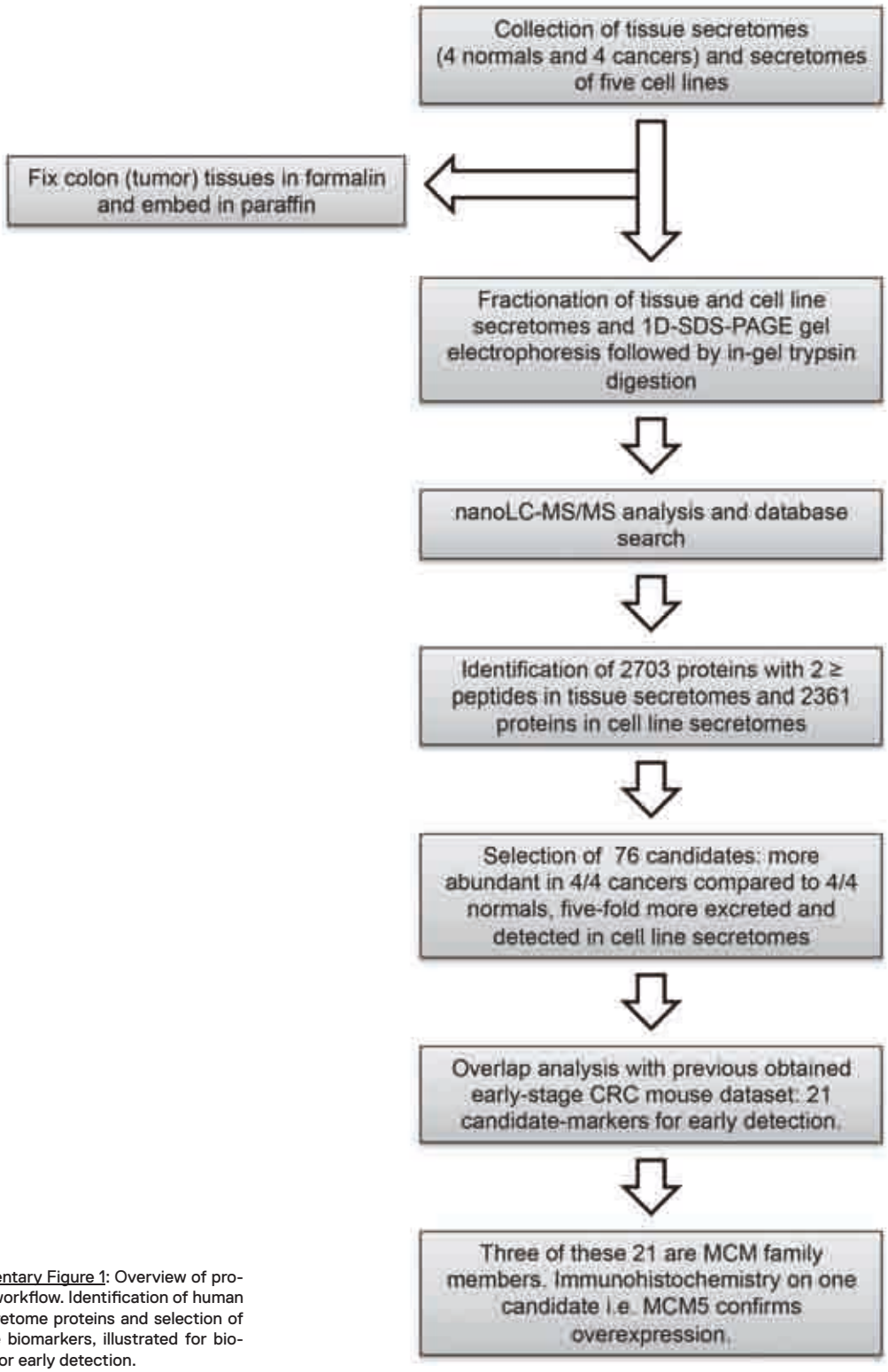
Auto-antibodies directed against tumor-associated or aberrantly-located antigens have been described in CRC and the presence of such antibodies suggests that the antigens were excreted by the tumor in order to initiate an immune response. A signature of 18 autoantibodies capable of distinguishing CRC patient serum from normal control serum contained antibodies against two proteins out of the list of 76 CRC candidate biomarkers presented in this study, namely TRIM28 and HNRPDL[32]. Interestingly, both of these proteins are annotated as nuclear, which further supports the notion that nuclear proteins are excreted and can end up in blood or stool.

Validation of secreted biomarker proteins in body fluids like blood or stool is crucial for the development of clinical applications. At present, this validation is hampered by limited availability of sensitive ELISA assays for the majority of candidate biomarker proteins. In the current study we therefore applied immunohistochemistry for verification of overexpression of the nuclear MCM5 protein in colon cancer, although this method is not suited to assess actual secretion. The finding of proteins that were previously validated in serum or stool as biomarkers for CRC (e.g. MCM proteins, TIMP-1 and LCN2) confirms that this approach is valid and warrants further validation of novel candidate markers[22,23,26]. New technologies such as peptide-capture antibodies in combination with SRM-MS and large-scale antibody-based plasma profiling as is currently developed by the Human Protein Atlas may provide better opportunities for future large-scale validation[33,34].

In conclusion, proteomic analysis of normal human colon and CRC tissue secretomes combined with CRC cell line secretome analysis is a powerful strategy to discover candidate protein biomarkers for CRC. These biomarkers have the potential to be used for development of blood-based or stool-based tests to support clinical management of CRC, although further studies in specified clinical settings are required to validate their clinical applicability.

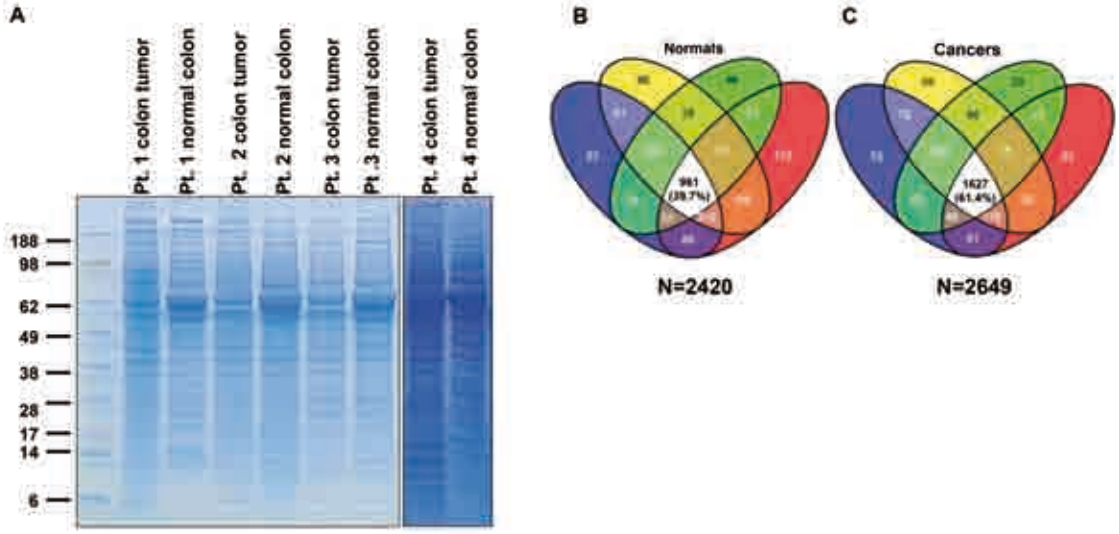
1 Parkin DM, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. CA: a Cancer Journal for Clinicians 2005;55:74–108.
2 Bosch LJW, Carvalho B, Fijneman RJA, Jimenez CR, Pinedo HM, van Engeland M, Meijer GA. Molecular tests for colorectal cancer screening. Clinical Colorectal Cancer 2011;10:8–23.
3 Karl J, Wild N, Tacke M, Andres H, Garczarek U, Rollinger W, Zolg W. Improved diagnosis of colorectal cancer using a combination of fecal occult blood and novel fecal protein markers. Clin. Gastroenterol. Hepatol. 2008;6:1122–8.
4 Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, Somerfield MR, Hayes DF, Bast RC, ASCO. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. Journal of Clinical Oncology 2006;24:5313–27.
5 Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol. Cell Proteomics 2002;1:845–67.
6 Celis JE. Proteomic Characterization of the Interstitial Fluid Perfusing the Breast Tumor Microenvironment: A Novel Resource for Biomarker and Therapeutic Target Discovery. Mol. Cell Proteomics 2004;3:327–44.
7 Fijneman RJA, de Wit M, Pourghiasian M, Piersma SR, Pham TV, Warmoes MO, Lavaei M, Piso C, Smit F, Delis-van Diemen PM, van Turenhout ST, Terhaar Sive Droste JS, Mulder CJJ, Blankenstein MA, Robanus-Maandag EC, Smits R, Fodde R, van Hinsbergh VWM, Meijer GA, Jimenez CR. Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diagnosis of human colorectal cancer. Clin. Cancer Res. 2012;18:2613–24.
8 Hoskins ER, Hood BL, Sun M, Krivak TC, Edwards RP, Conrads TP. Proteomic analysis of ovarian cancer proximal fluids: validation of elevated peroxiredoxin 1 in patient peripheral circulation. PLoS ONE 2011;6:e25056.
9 Teng P-N, Hood BL, Sun M, Dhir R, Conrads TP. Differential proteomic analysis of renal cell carcinoma tissue interstitial fluid. J. Proteome Res. 2011;10:1333–42.
10 Societies DFOBS. Code for proper secondary use of human tissue in the Netherlands. www.federa.org; n.d.
11 de Wit M, Jimenez CR, Carvalho B, Belien JAM, Delis-van Diemen PM, Mongera S, Piersma SR, Vikas M, Navani S, Pontén F, Meijer GA, Fijneman RJA. Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. Gut 2012;61:855–64.
12 Piersma SR, Fiedler U, Span S, Lingnau A, Pham TV, Hoffmann S, Kubbutat MHG, Jimenez CR. Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. J. Proteome Res. 2010;9:1913–22.
13 Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 2003;75:4646–58.
14 Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 2002;74:5383–92.
15 Pham TV, Jimenez CR. An accurate paired sample test for count data. Bioinformatics 2012;28:i596–i602.
16 Pham TV, Piersma SR, Oudgenoeg G, Jimenez CR. Label-free mass spectrometry-based proteomics for biomarker discovery and validation. Expert Rev. Mol. Diagn. 2012;12:343–59.
17 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, Mering von C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39:D561–8.
18 Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 2011;27:431–2.
19 Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005;21:3448–9.
20 Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Meth 2012;9:471–2.
21 Bendtsen JD, Jensen LJ, Blom N, Heijne Von G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. 2004;17:349–56.
22 Davies RJ, Freeman A, Morris LS, Bingham S, Dilworth S, Scott I, Laskey RA, Miller R, Coleman N. Analysis of minichromosome maintenance proteins as a novel method for detection of colorectal cancer in stool. The Lancet 2002;359:1917–9.
23 Nielsen HJ, Brünner N, Jorgensen LN. Plasma TIMP-1 and CEA in detection of primary colorectal cancer: a prospective, population based study of 4509 high-risk individuals, Scandinavian Journal of Gastroenterology, Informa Healthcare. Scandinavian ... 2011.
24 Draht MX, Riedl RR, Niessen H, Carvalho B, Meijer GA, Herman JG, van Engeland M, Melotte V, Smits KM. Promoter CpG island methylation markers in colorectal cancer: the road ahead. Epigenomics 2012;4:179–94.
25 Zhu Y-M, Huang Q, Lin J, Hu Y, Chen J, Lai M-D. Expression of human DNA methyltransferase 1 in colorectal cancer tissues and their corresponding distant normal tissues. Int J Colorectal Dis 2007;22:661–6.
26 Sun Y, Yokoi K, Li H, Gao J, Hu L, Ben Liu, Chen K, Hamilton SR, Fan D, Sun B, Zhang W. NGAL Expression Is Elevated in Both Colorectal Adenoma–Carcinoma Sequence and Cancer Progression and Enhances Tumorigenesis in Xenograft Mouse Models 2011.
27 Findeisen P, RÖCKEL M, Nees M, RÖder C, Kienle P, Knebel Doeberitz M Von, H K, MM N. Systematic identification and validation of candidate genes for detection of circulating tumor cells in peripheral blood specimens of colorectal cancer patients. Int J Oncol 2008:1001–10.
28 Hlavata I, Mohelnikova-Duchonova B, Vaclavikova R, Liska V, Pitule P, Novak P, Bruha J, Vycital O, Holubec L, Treska V, Vodicka P, Soucek P. The role of ABC transporters in progression and clinical outcome of colorectal cancer. Mutagenesis 2012;27:187–96.
29 György B, Szabó TG, Pásztói M, Pál Z, Misják P. Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. Cellular and Molecular 2011.
30 Kharaziha P, Ceder S, Li Q, Panaretakis T. Tumor cell-derived exosomes: A message in a bottle. Biochim. Biophys. Acta 2012;1826:103–11.
31 Mathivanan S, Ji H, Simpson RJ. Exosomes: extracellular organelles important in intercellular communication. J Proteomics 2010.
32 Kijanka G, Hector S, Kay EW, Murray F, Cummins R, Murphy D, MacCraith BD, Prehn JHM, Kenny D. Human IgG antibody profiles differentiate between symptomatic patients with and without colorectal cancer. Gut 2009;59:69–78.
33 Quantification of Proteins Using Peptide Immunoaffinity Enrichment Coupled with Mass Spectrometry 2011.
34 Schwenk JM, Igel U, Neiman M, Langen H, Becker C, Bjartell A, Pontén F, Wiklund F, Grönberg H, Nilsson P, Uhlén M. Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays. Mol. Cell Proteomics 2010;9:2497–50.

SUPPLEMENTARY FIGURE 1



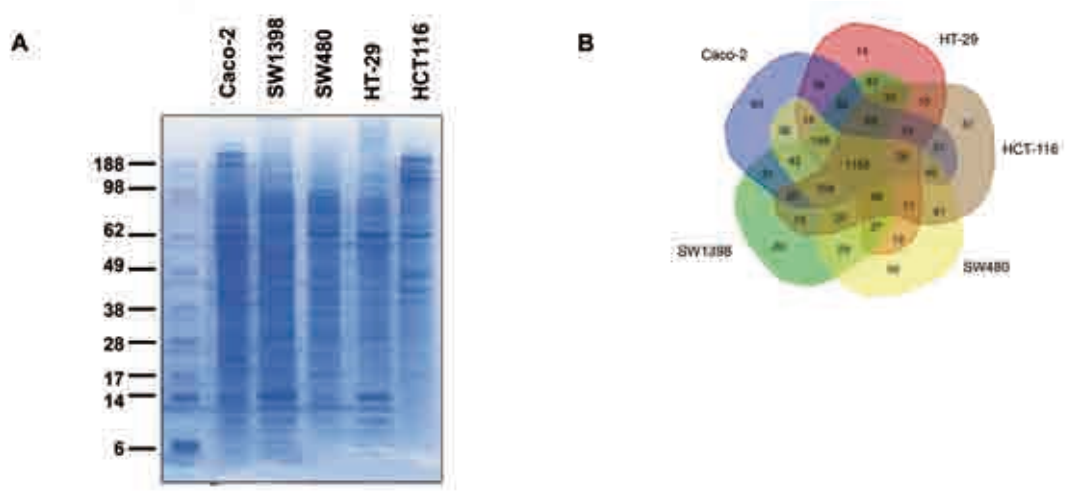
Supplementary Figure 1: Overview of proteomics workflow. Identification of human CRC secretome proteins and selection of candidate biomarkers, illustrated for biomarkers for early detection.

SUPPLEMENTARY FIGURE 2



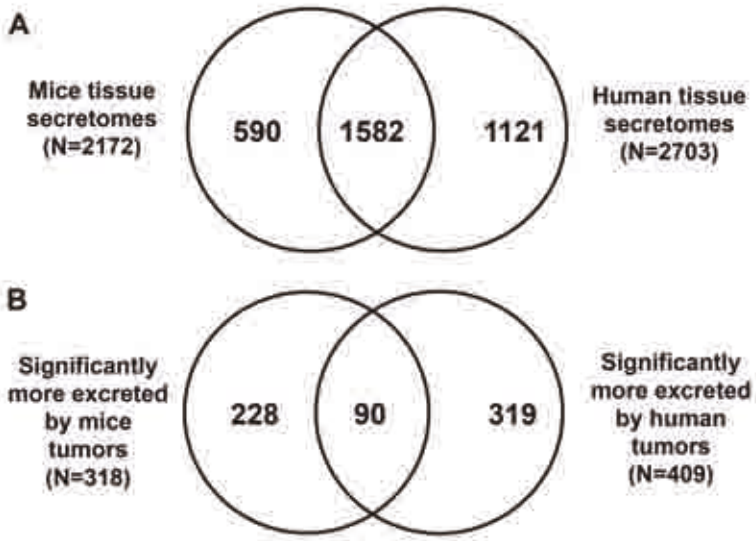
Supplementary Figure 2: Gel coomassie staining and protein identifications in tissue secretomes. A: cancer secretomes and paired normal colon tissue secretomes of four patients (Pt.). B,C: Venn diagrams illustrate overlapping and unique protein identifications. B: normal tissue secretomes. C: tumor tissue secretomes. A total of 2420 different proteins were identified in all normal colon tissue secretomes, of which 961 in all four samples. In the tumor secretomes 2649 proteins were identified of which 1627 in all four samples. Color code: patient one in blue, patient two in yellow, patient three in green and patient four in red.

SUPPLEMENTARY FIGURE 3



Supplementary Figure 3: Gel coomassie staining and protein identifications in CRC cell line secretomes. A: Secretomes of colon cancer cell lines Caco-2, SW1398, SW480, HT-29 and HCT 116. B: Venn diagram illustrates overlapping and unique protein identifications. A total of 2361 different proteins were identified in all CRC cell line secretomes, of which 1158 (49%) in all five samples.

SUPPLEMENTARY FIGURE 4



Supplementary figure 4: Proteins excreted by mouse and human colon tumors. Venn diagram illustrates overlapping and unique protein identifications among mouse and human colon tumor datasets. A: total protein identifications in mouse colon (tumor) tissue secretomes [7] and in human colon (cancer) tissue secretomes (current study). In total 1580 different proteins were commonly identified in both studies. B: All proteins previously found to be significantly more excreted by mouse colon tumors and proteins significantly more excreted by human CRC tissue, with 90 proteins that overlap.